Method for Identifying Organisms by Means of Comparative Genetic Analysis and Primers and Hybridization Probes for Carrying out this Method

Description

The present invention relates to a method for the genetic analysis of organisms of different species of animal and/or plant by studying coding and non-coding areas of highly conserved genes or pseudogenes and their homologues with various species of animal and plant.

The known methods serving for simply, readily and precisely determining gene sequences to detect relationships and to identify organisms are based on the use of oligonucleotides which are specific to a species. These oligonucleotides are necessary to provide sufficient genetic material for the subsequent sequence reaction by means of polymerase chain reaction or other methods. An oligonucleotide is usually a short synthetically produced molecule which binds to a specific gene segment and has a specific sequence complementary to this strand. The drawback of methods based on the use of oligonucleotides consists in that gene sequences of different species usually differ strongly. In order to sequence organisms of different species, species-specific DNA sequences must, as a rule, be determined by means of methods which are costly and time-consuming, and the corresponding oligonucleotides which bind specifically to these DNA sequences must be synthesized in a second step. These two steps are usually required for each organism to be studied of unknown origin. For a subsequent analysis for determining the identity or the relationship due to the gene sequence the DNA of the organisms to be studied must then be tested with a usually large number of different oligonucleotides. Since usually only oligonucleotides of few

species are available, the analysis often fails to be a success in the case of rare species of animal. The analysis is also time-consuming and expensive.

Other methods which utilize e.g. restriction length polymorphisms of the organisms or methods used for amplifying a mixture of short oligonucleotides randomized as regards their sequence (random amplified polymorphic DNA, RAPD) do not result in an accurate and unambiguous analysis of the gene sequence of individual animals. As a result, it is difficult to determine the relationship between organisms by means of these methods.

It is thus the object of this invention to provide a method for the genetic analysis of organisms, which is highly sensitive, supplies reliable results while consuming little time, is also suited for major serial examinations and routine tests, and can optionally also be carried out automatically. The object of this invention also consists in developing means or products for carrying out this method.

The invention is realized according to the claims, the subclaims being preferred variants.

The subject matter of the invention relates to a method by which the gene sequence of a gene area of various species can be determined readily, simply and safely in a reproducible way. Three preconditions must be met here: The oligonucleotides serving as primers for the amplification of the DNA must bind to areas of the genome which are highly conserved to ensure an amplification of genetic material by means of an identical oligonucleotide pair in all or the greatest possible number of different organisms. These oligonucleotides must cover an area which has a great sequence diversity between different species to enable a differentiation. The area which is covered by the oligonucleotides used as primers should be as small as possible to maximize the yields of amplificates and to ensure that copies can also be obtained from strongly degraded DNA in the starting material.

According to the invention this object is achieved by providing a method of determining the identity or the

relationship by comparing coding and non-coding areas of highly conserved genes of pseudogenes and homologues. As a result, it is ensured that a single oligonucleotide pair binds to DNA sequences highly conserved between various species and thus enables a gene segment identical for all species to be amplified. The oligonucleotides comprise one or more gene areas having the greatest possible sequence differences between different species. The determination of the gene sequence of this highly polymorphous gene area in a subsequent reaction step enables the gene sequence to be allocated to a specific species.

Genotyping is made by sequencing or by other methods which are suited for the detection of sequence variants. This comprises genotyping methods assisted by polymerase chain reaction (PCR), such as allele-specific PCR, other genotyping methods using oligonucleotides (e.g. "dot blotting", or "Oligonucleotide Ligation Assays" (OLA)), methods using restriction enzymes, analysis of length polymorphisms and "single nucleotide polymorphisms" (SNP), analysis by means of spectroscopic methods such as "matrix-assisted laser desorption/ionization mass spectroscopy" (MALDI), chromatographic methods such as DHPLC for separating DNA strands of differing lengths and sequences and in principle any method available at present or in the future for variant detection, including DNA, RNA and PNA hybridization methods, light cycler technology, TaqMan and molecular beacon technology and the chip technology in all its technological realizations.

The following steps are preferably carried out in the method according to the invention:

a) DNA isolation: DNA is isolated and purified from blood samples, tissues, hair, foodstuffs and samples containing DNA.

b) Polymerase chain reaction: The polymerase chain reaction serves for amplifying DNA for the subsequent sequencing reaction. In the polymerase chain reaction, one or more oligonucleotide pairs bind to the DNA to be analyzed (template DNA) of genes which are highly conserved between organisms of various species. In each case, one of the

molecules of an oligonucleotide pair (sense and antisense oligonucleotide) is complementary to one of the two template DNA strands at the 5' or 3' end of a DNA sequence. The binding is oriented such that the synthesis products obtained in an oligonucleotide-controlled polymerase chain reaction using one of the two oligonucleotides each may serve, following denaturation, as a matrix for binding the respectively other oligonucleotide. The oligonucleotide pair flanks the area which shall be copied. These oligonucleotides are extended in accordance with the nucleotide sequence of the template strand by means of polymerase and following the addition of nucleotide building blocks. Binding, extension and denaturation take place at different temperatures and are usually carried out 20 to 35 times in succession so as to multiply exponentially the area covered by the oligonucleotides.

c) Agarose gel electrophoresis: The DNA fragments are separated from the oligonucleotides in an agarose gel with a voltage being applied, and the band specific to the PCR product is excised using a scalpel and purified.

d) Sequencing reaction: Another polymerase is added as well as all nucleotide building blocks and special nucleotide building blocks which terminate the chain in the extending reaction. As a result, DNA fragments form which differ in length by one nucleotide each. Areas within the gene are sequenced which as regards their gene sequence are polymorphous between the different species. Therefore, the DNA sequence characteristic of a species can be determined and allocated to a species.

e) Polyacrylamide gel eletrophoresis: If after concluding the sequencing reactions the differently long DNA strands are separated on a high-resolution polyacrylamide gel in an electric field, shorter DNA strands will migrate more rapidly than longer strands. In the pattern of bands forming, the order from shorter band to the respective band next in length - allocated to the corresponding bases A, C, G or T - corresponds to the complementary DNA sequence of the template. As a result, the sequence of the DNA strand, which was amplified by means of polymerase chain reaction

beforehand, becomes readable.

f) Comparative analysis of the gene sequence between the animal species and storage of the sequencing data.

For the amplification the method according to the invention preferably uses oligonucleotide pairs which bind to coding or non-coding areas of highly conserved genes or pseudogenes and their homologues, i.e. to genes or pseudogenes and their homologues which show no, or only minor, sequence differences between the individual species. This ensures that an amplificate can be formed in each gene segment to be analyzed of the most differing species by means of a primer pair.

The method according to the invention preferably serves for analyzing areas of the gene or pseudogene and their homologues, which differ as regards their gene sequence between organisms of different species. These may be sequences of coding or non-coding DNA.

The method according to the invention uses for certain batches for the polymerase chain reaction several differently long oligonucleotide pairs in a reaction mixture (multiplex PCR), all of which are partially identical with a starting sequence. Here, the sense or antisense oligonucleotides differ each as regards the nucleotide sequence such that the lengths of some oligonucleotides differ in the 3' region by one or more nucleotides. The use of several oligonucleotides differing in length shall ensure that binding of the oligonucleotides to the template DNA will be possible even if the template DNA differs from some oligonucleotides used as regards the 3' region. The correspondence in the 3' region of the oligonucleotides with the template DNA is essential for the specific amplification and should thus be as precise as possible in this area.

Example:

sense:

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gac -3',

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat -3',

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat t -3',

antisense:

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt c -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta c -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc t -3'.

In another preferred embodiment of the method according to the invention, multiple oligonucleotide pairs are used in a multiplex PCR, which have equal length but differ at one or more positions of the 3' end of the oligonucleotides as regards their nucleotide sequence. The use of a reaction mixture of several oligonucleotides differing as regards the base sequence at the 3' end of the oligonucleotide shall ensure that an amplificate can be formed in the most different species by means of a primer pair.

The method according to the invention preferably analyzes segments of the gene or pseudogene and their homologues which differ as regards their gene sequence between organisms of different species. These may be sequences of coding or non-coding DNA.

The method according to the invention uses for certain batches for the polymerase chain reaction several differently long oligonucleotide pairs in a reaction mixture (multiplex PCR), all of which are partially identical with a starting sequence. Here, the sense or antisense oligonucleotides differ each as regards the nucleotide sequence such that the lengths of some oligonucleotides differ in the 3' region by one or more nucleotides. The use of several oligonucleotides differing in length shall ensure that binding of the oligonucleotides to the template DNA will be possible even if the template DNA differs from some oligonucleotides used as regards the 3' region. The correspondence in the 3' region of the oligonucleotides with the template DNA is essential for the specific amplification and should thus be as precise as possible in this area.

Example

sense:

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gac -3',

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat

-3',

5' - cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat t -3',

antisense:

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt c -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta c -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc t -3'.

In another preferred embodiment, the method according to the invention uses multiple oligonucleotide pairs in a multiplex PCR, which have equal length but differ at one or more positions of the 3' end of the oligonucleotides as regards the nucleotide sequence. The use of a reaction mixture of several oligonucleotides differing as regards the 3' end of the oligonucleotide shall ensure that the oligonucleotides are bound to the template DNA even if the template DNA differs at the 3' binding site of the oligonucleotide from the usually used oligonucleotide. For this purpose, a mixture of different oligonucleotides which have all conceivable nucleotide sequences at their 3' end is provided for the amplification.

Example:

sense:

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gaa -3',

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gac -3',

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gag -3',

5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gat -3',

antisense:

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt a -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt c -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt g -3',

5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt t -3'.

When suitable oligonucleotide sequences are selected, the method according to the invention pays attention to the fact that as many oligonucleotides as possible correspond at their 3' end with the first nucleotide of the codon which codes for

the highly conserved amino acid. It follows from theoretical considerations and also on the basis of observations that every second nucleotide has within a codon a degree of correspondence between organisms of differing species higher than that of the first or third nucleotide of the codon. Since the amplification will usually only function if along with other preconditions the nucleotide located at the 3' end of the oligonucleotide is exactly complementary to the opposite nucleotide of the template DNA strand - i.e. A faces T or G faces C -, the sequence of the oligonucleotides employed is chosen such that the last nucleotide at the 3' end of the oligonucleotide binds, if possible, to the most highly conserved nucleotide of the codon coding for an amino acid.

The method according to the invention preferably uses for different segments of the gene, of the pseudogenes and its homologues a single oligonucleotide pair each, which binds to the highly conserved tumor suppressor gene PTEN/MMAC1, its pseudogene and their homologues in different species, namely in the area of the gene which shows major correspondence between the species [Steck *et al.*, 1997]. This area concerns the entire coding sequence of the gene, pseudogene and its homologues as well as exon-intron transitions and 5'- and 3'-untranslated regions of the gene, pseudogene and their homologues.

The areas amplified by means of the above described methods comprise sequence segments which have more or less great sequence differences between the individual species. This applies in particular to intron regions and specifically to intron 4. The oligonucleotide pair used for the amplification may bind to exons 4 and 5, since the PTEN/MMAC1 gene, pseudogene and their homologues show no, or only slight, differences between the species in exon 4 and exon 5. The intron region 4 covered by both oligonucleotides, which in contrast to the exon regions 4 and 5 between the species has considerably greater sequence differences, is amplified. Thereafter, part of this intron region is sequenced preferably by means of a sequencing reaction (see figure 1).

By means of comparative analyses, it is possible to determine due to the correspondence of intron sequences whether different samples belong to an identical species. It is also possible to determine the relationship of different species by means of comparative analysis of an intron region comprising only some to several hundred bases of the PTEN/MMAC1 gene and its homologues due to the similarity of the sequences. The general rule is that for the differentiation of closely related species the intron segments which must be studied have to be longer than those of distantly related species. This procedure described for exons 4/5 and intron 4 may also be applied to all of the other introns enclosed by exons. It also applies to pure exon regions, pseudogenes and the 5'- and 3'-untranslated regions as well as their homologues, in these cases the sequence differences between the individual species being less than in the intron regions.

In a preferred embodiment of the method according to the invention, highly conserved pseudogenes and their homologues are studied in various organisms and used for determining the species-specific gene sequence. These pseudogenes and their homologues have the advantage that they are also highly conserved as regards their nucleotide sequence and allow an amplification in certain species. However, the sequence differs in some areas of the pseudogene and its homologues between different species, so that pseudogenes and their homologues can be used for the species-specific characterization of organisms. Since the intron regions lack, they - like pure exon regions and the 5'- and 3'-untranslated regions - are suitable for the analysis, in particular of DNA degraded due to environmental influences, because of the small size.

An advantageous execution of the method utilizes the deletion identified by inventors and having a length of 9 base pairs in a PCR product (see figure 3 and Example 2), which corresponds to a gene sequence variant of PTEN/MMAC1 and was amplified from DNA of pig cells. This deletion having a

length of 9 base pairs is typically found in domestic pigs and all examined wild boars and as for the rest in no other examined species. In a variant of an embodiment of the invention this difference in length serves for proving pig meat in foodstuffs. This variant in length is genotyped by sequencing or by other methods suitable for detecting this deletion. They comprise PCR-assisted genotyping methods such as PCR by means of species-specific oligonucleotides, hybridization techniques such as the light cycler technology or other genotyping methods using restriction enzymes, and in principle any method available at present or in the future for detecting variants, including the chip technology and all of their technological realizations. Deletions and insertions, as found in different species in intron 4, exon 8 and in the 5'-untranslated region, can also be used correspondingly for identifying species (see annex: "List of species sequences").

The method according to the invention was made with DNA from different species as a model system and the gene sequence in the area of two segments in intron 4 of the PTEN/MMAC1 gene and its homologues was determined. It was possible to amplify all species with only one oligonucleotide pair. In the sequencing reaction with only one oligonucleotide and in the subsequent analytical polyacrylamide gel electrophoreses it turned out that all investigated species differ as regards the nucleotide sequence in the intron region (see annex: "List of species sequences"). Since the PTEN/MMAC1 tumor suppressor gene and its pseudogene and their homologues are conserved, the method can in other species also amplify successfully the corresponding gene sequences of lower organisms and possibly plants by means of an oligonucleotide pair. The method according to the invention is thus suited to determine the identity and relationship of various organisms. A data library was established which can be used for identifying different species and humans. It comprises PCR primers, sequencing primers and hybridization probes as well as sequences of the coding and non-coding areas including select highly variant intron regions, of exon regions, of the

5'-untranslated region of the gene and its homologues as well as of the pseudogene and its homologues from the most differing vertebrates (see annex: "List of species sequences"). On the one hand, the method according to the invention is suited to determine readily, simply and safely in a reproducible manner the relationship in certain species which are clearly classified (phylogenetic analyses). On the other hand, it is possible to determine by a comparison with gene sequences which can clearly be allocated to a species, the identity of tissue samples, blood samples and foodstuffs and all samples which contain DNA and are of unknown origin. Therefore, the method is also suited for applications in forensic medicine. Since in a preferred embodiment of the method according to the invention DNA is used as the starting material, organic samples (e.g. blood, saliva, tissue residues) can clearly be allocated to human or animal origin because of their gene sequence. Due to the possibility of comparing the collected DNA sequence with an already established data library with DNA sequences of known species it is possible to make statements on the species. If the gene sequence is unknown, it is possible due to sequence similarities to make statements on the relationship which the species to be studied has with a comparative DNA.

An advantageous embodiment of the invention utilizes hybridization probes for distinguishing the DNAs of various species among one another. This method uses differing ways for distinguishing the species on the basis of the LightCycler analysis system (company of Roche Molecular Biochemicals) using hybridization probes (patents WO 97/46714; WO 97/39008).
The LightCycler analysis system makes possible the amplification, detection and specific analysis of DNA of differing species and their differentiation within the shortest possible time and with moderate expenditure. The LightCycler is a micro-volume fluorimeter having a thermocycler combining rapid thermocycling with real-time fluorescence observation during the PCR (Wittwer *et al.*, 1997a). By means of this technology, the time for the

amplification and the detection of nucleic acids is reduced from about 5 hours to about 30 minutes. For a specific detection during the PCR reaction, the synthesis can be observed on the basis of the fluorescence resonance energy transfer (FRET) via two adjacent hybridization probes labeled with fluorescent dyes. Here, one probe is labeled at its 3' end with a donor fluorophore (usually fluorescein) and the adjacent probe is labeled at its 5' end with an acceptor fluorophore. During FRET, the donor dye is excited by an external light source and emits light which is absorbed by the acceptor fluororphore. The latter in turn emits light having another wavelength which is measured specifically. This FRET can only be made if both probes hybridize side by side within the amplification pair at a distance of about 1-5 bp, on the target DNA. Here, the sequence of the probes is selected such that it is complementary to the target area [Wittwer *et al.* 1997b; Lay, 1997, Bernard, 1998; Ririe, 1997; Bernard, 1999; Nauck, 1999a; Nauck, 1999b, Kreuzer, 1999; Kyger, 1998; Mangasser-Stephan, 1999; Aslandis, 1999].

In order to detect the specific amplification product and to differentiate differences in the target sequence, the LightCycler provides the possibility of carrying out melting point analyses. The melting point analysis is based on the fact that two complementary DNA strands are separated at a characteristic temperature, the melting temperature, into two individual strands. This melting temperature depends on the base composition in which sequences rich in GC have a higher melting point than sequences with predominantly AT bases. If a melting point analysis is carried out by removing more or less well complementary hybridization probes from the template DNA under certain temperature conditions, there is a melting profile characteristic of the fitting between probe and target sequence to be analyzed in the form of a fluorescence intensity measurement. As a result, the changes in the target sequence can be detected. If the sequences are fully complementary to one another, the probes will fully hybridize. If the target sequence contains changes in its base sequence, the melting point of the probes is lowered

correspondingly. Due to a continuous detection of the fluorescence up to the melting of the probes, a melting curve can be prepared for each sample in the form of the fluorescence intensity as a function of the temperature. A comparison of the melting peaks made by the first negative derivative of the fluorescence to the temperature (-dF/dt vs T) enables a verification and differentiation of various DNAs. Fragments having a lower or higher melting temperature can be distinguished clearly.

In particularly preferred variants of embodiments of the invention, oligonucleotide pairs were found for the polymerase chain reaction (PCR) which bind to the highly conserved tumor suppressor gene PTEN/MMAC1, its pseudogene and their homologues in areas which show large correspondence between the species. These are usually exon regions or untranslated areas of the 3' and 5' ends of the gene, its pseudogene and their homologues.

These oligonucleotide pairs allow to amplify small gene segments containing areas which have differing base sequences between the individual species, including base substitutions, base deletions and base insertions. Examples 4, 5 and 6 use a 9-base pair deletion of the pig homologue of the PTEN/MMAC1 pseudogene and numerous further sequence variants of the gene, pseudogene and their homologues in various species to produce by means of specific hybridization probes differing melting profiles with differing species, which clearly distinguish the species from one another. This preferred area of Examples 4 to 6 relates exclusively to exon 5 of the gene, pseudogene and their homologues. The PCR primers used for the amplification of the corresponding gene segment read as follows:

Sense primer: PTEN se 5'- atc ttg acc aat ggc taa gtg -3'
Antisense primer: Zoo44aRV 5'- ttgt ctc tgg tcc tta ctt c -3'

Hybridization probes according to the invention were selected on the one hand with the target area of the 9-base pair deletion of the pig pseudogene. Probe A1/A2 enables a

distinction of the pig DNA from all of the other species.

Since by the above PCR primer pair, it is not only the pseudogene but also the gene area of exon 5 of the pig homologue of PTEN/MMAC1 that is amplified, another probe pair was provided with the complementary sequence of the pig homologue. This is probe C1/C2.

Since the different species in this gene area show minor sequence differences which should be used for a detailed differentiation among one another, a third probe pair which corresponds to the sequence of the human pseudogene in the selected area was constructed (probe B1/B2).

According to the invention, the probes A1/A2 are concerned:
specific to PTEN pseudogene pig:
A1: 5'- tgc ata ttt gtt tca tcc ggg caa att - fluorescein -3'
A2: 5'- LC Red 705 - tta aag gca caa gat ttc tat ggg ga - ph -3'

Probes B1/B2: specific to PTEN pseudogene man:
B1: 5'- tgc ata ttt att aca tcg ggg caa att - fluorescein -3'
B2: 5'- LC Red 640 - aag gca caa gag gcc cta gat ttc ta - ph -3'

Probes C1/C2: specific to PTEN homologue pig:
C1: 5'- tgc ata ttt gtt aca tcg ggg taa att - fluorescein -3'
C2: corresponds to probe B2

The positions of probes A1, A2, B1, B2, C1 and C2 in exon 5 are shown in figure 4.

The separate use of these three probe pairs and the use of the probes in various possible combinations (1 donor dye + 1 acceptor dye) yields for each individual species a characteristic panel of differing melting points enabling a clear distinction between the species (see figures 5 and 7).

The use of these hybridization probe combinations also permits to carry out studies in reaction mixtures of two or more different species.

A parallel analysis/detection of the fluorescence of two different wavelengths (640 nm and 705 nm) is possible in a multiplex reaction with two different probe pairs such that the respective donor probes are labeled differently. The acceptor probes may be identical or differ as regards their sequence. Having concluded the melting point analysis, a melting point specific to the corresponding probe and the target sequence covered by it is obtained for each of the two wavelengths.

The alternative combination from two different donor probes and one acceptor probe of one wavelength yields for a reaction mixture of two different species having minor sequence differences in the target area two melting points within one wavelength.

In general, the following general multiplex reaction batches are possible which are characterized in that at least one hybridization probe pair is used and at least one gene segment is amplified, differing hybridization probe pairs hybridize to differing gene segments, and the melting points of the different combinations are determined and compiled for each species into a panel or used for the identification.
These general multiplex reaction batches can also be characterized in that not only at least one hybridization probe pair is used and not only at least one gene segment is amplified but also DNA of at least one species is used and thus different hybridization probe pairs hybridize to differing gene segments of different species, and the melting points of the different combinations are determined and compiled for each species into a panel or used for the identification.

On this basis, the following analytical approaches are possible, melting point overlaps having to be avoided.

a) Analysis of an unknown species sample with two different donor probes and one acceptor probe of one wavelength or of differing wavelengths,
b) Analysis of two or more unknown species samples with two different donor probes and one acceptor probes of one wavelength or of different wavelengths,
c) analysis of an unknown species sample with a mixture of two or more donor probes and acceptor probes each,
d) analysis of a mixture of unknown species samples with one donor probe and one acceptor probe,
e) analysis of a mixture of unknown species samples with two or more different donor probes and two or more acceptor probes of one or more different wavelengths.

This procedure can be applied to all highly conserved segments of the PTEN/MMAC1 gene, pseudogene and their homologues and also to other highly conserved genes when species shall be distinguished.

Select primers for the amplification are defined in claims 18 to 25 and select hybridization probe sequences for the LightCycler application are defined in claims 34 to 42 for different exons of the PTEN/MMAC1 gene, pseudogene and their homologues. Due to the high conservation of the gene in the evolution, primers and hybridzation probes are also possible in all of the other conceivable areas of the gene, pseudogene and their homologues which can be used according to the above described principle for distinguishing species.

These probes and/or primers can be combined as desired in accordance with the multiplex principle with the aim of differentiating species.

The general rule is that all of the described methods and all methods conceivable at present and in the future for analyzing the DNA sequence variants can be applied to both the sense strand and antisense strand. This equivalence principle also applies to all PCR, sequence primer and hybridization probes which are described in this patent

application and the described gene sequences of the individual species of animal.

The following examples shall explain the invention in more detail.

Example 1

In this experiment, the sequence diversity between human DNA and elephant DNA is determined. The arrangement shown in figure 1 was chosen. The comparative sequence analysis was carried out by determining the species-specific sequences of intron 4 of the tumor suppressor gene PTEN/MMAC1 and its homologues. An oligonucleotide pair which is specific to areas within exons 4 and 5 of the DNA sequence of this gene for mice was chosen and subsequently synthesized (company of Amersham Pharmacia).

Oligonucleotide 1: 5'- cga cgt tgt aaa acg acg gcc agt tgt gct gag aga cat tat gac -3'
Oligonucleotide 2: 5'- cag gaa aca gct atg act tgt ctc tgg tcc tta ctt c -3'

The oligonucleotides for the polymerase chain reaction were constructed such that they have one sequence each at their 5' end (underlined) which is complementary to the two oligonucleotide used for the sequencing reaction. As a result, it is ensured that the specific oligonucleotide parts bind in each case specifically to the template DNA (not underlined) but the oligonucleotides used for sequencing may differ from those used for the polymerase chain reaction, which adds quite generally to the quality of the sequencing reaction. The specific portion of the oligonucleotides binds in each case to exon 4 and exon 5 of the PTEN/MMAC1 tumor suppressor gene and its homologues.

A veterinarian provided 3 ml of residual blood of an African elephant resulting from a routine operation. The DNAs from elephant blood and from human blood were isolated by means of the QIAgen kit (QIAGen company) in accordance with the protocols from the manufacturer and stored at -20°C until they were used.
In order to generate sufficient amounts of DNA for the sequencing reaction, a polymerase chain reaction (PCR) was carried out. 3 sample batches were provided and the following

volumes and final concentrations or amounts of substrates and units of polymerase were used: reaction vessel 1: elephant DNA 1 µl (50 ng); reaction vessel 2: human DNA 1 µl (50 ng); reaction vessel 3: no DNA but 1 µl distilled water instead (negative control). All sample batches were provided with the following: 14.35 µl distilled water; dNTPs (Promega company): 4 µl (200 µl); $MgCl_2$ (InViTek): 1 µl (2 mM); 10 x buffer (InViTek) consisting of 160 mM $(NH_4)_2SO_4$, 500 mM Tris-HCl, pH 8.8, 0.1 % Tween 20: 2.5 µl; oligonucleotides 1 and 2: 1 µl (0.2 µM); Taq polymerase (InViTek): 0.15 µl (0.75 units).
All in all, 35 cycles were carried out in a Perkin Elmer Thermocycler 9600, denaturing taking place at 94°C for 50 seconds each, the oligonucleotides binding to the DNA strands at 53°C for 50 second, and extension taking place at 72°C for 60 seconds (one second longer per cycle). Before the first cycle was started, denaturing was carried out at 94°C for 3 minutes, and a last extension phase of 10 minutes at 72°C was carried out after the last cycle.
In order to check the success of the PCR, to isolate the amplified DNA and check the negative control, an agarose gel electrophoresis was carried out: A 0.8 % agarose gel was produced by adding ethidium bromide, and 1 x TAE buffer was added. 13 µl PCR product each was added by pipetting to prepared sample bags or pockets to the gel matrix by adding 2 µl dye and a voltage of 100 V was applied. Electrophoresis was stopped after 20 to 30 minutes, the DNA bands were made visible under a U.V. lamp and thereafter excised using a scalpel. The purification of the PCR product was possible by placing the excised bands on MicroSpin columns (Amersham Pharmacia) and centrifugation at 1020 g for 10 minutes. The eluate was diluted depending on the strength of the band on the agarose gel with up to 30 µl distilled water.

The purified PCR products were subjected to a cycle sequencing reaction using oligonucleotides having the sequence 5' - Cy-5- cag gaa aca gct atg ac -3'. The oligonucleotides were labeled at the 5' end with the dye Cy-5. 4 reaction vessels (A, C, G, T) were provided per PCR

product, and the following volumes and concentrations were chosen: 3 µl PCR product each per A, C, G, T; 1 µl each per A, C, G, T reagent (Amersham Pharmacia) consisting of Tris-HCl (pH 9.5), $MgCl_2$, Tween 20, Nonidet P-40, 2-mercaptoethanol, dATP, dCTP, 7-deaza-dGTP, dTTP, thermoresistant pyrophosphatase and thermo sequenase DNA polymerase, reaction vessels A, C, G and T containing correspondingly ddATP, ddCTG, ddGTP and ddTTP. Thereafter, 1 µl oligonucleotide of the above mentioned sequence (0.5 µM) is added. The following reaction conditions were chosen: All in all, 25 cycles were carried out in a Perkin Elmer Thermocycler 9600 (Perkin Elmer company), denaturing taking place at 94°C for 20 seconds each, the oligonucleotides binding to the amplificate at 54°C for 30 seconds, and the DNA strands being extended while raising the temperature to the denaturing temperature. Prior to the commencement of the first cycle, denaturing was carried out at 94°C for 3 minutes and 30 seconds, and the last extension phase of 5 minutes was carried out at 72°C after the last cycle.

In order to make visible the sequence, the products were separated on a polyacrylamide gel in an electric field. For this purpose, the automatic laser fluorescence detection system A.L.F.express from the company of Amersham Pharmacia was chosen. The gel matrix was composed of 16.8 g urea (Gibco BRL), 5.2 ml 50 % long ranger gel solution (FMC company) and 4 ml 10 x TBE (Gibco BRL), which were diluted using distilled water to give 40 ml. The gel was polymerized by adding 140 µl 10 % APS (Merck company) and 20 µl TEMED (Serva company). 5 µl formamide loading dye were added to each reaction vessel A, C, G, T, and added by pipetting to the prepared sample pockets. The following electrophoresis conditions were chosen: 1000 V, 40 mA, 40 W.

Following the gel electrophoresis, the sequences were generated by the A.L.F.express system (Amersham Pharmacia) and could then be compared with one another. Figure 2 shows the determined nucleotide sequence differences. A comparison of the sequences between elephant DNA and human DNA in the exon region yielded a great correspondence and served as a

control for the PTEN/MMAC1 specificity, whereas the intron region differed greatly.

Example 2

In this experiment, the sequence of a pig liver provided by a butcher's is compared with already available DNA of a pig. Beef salami served as a control. DNAs were obtained from 50 to 60 mg pig liver and 50 to 60 mg beef salami using the QIAgen kit (QIAGen) and purified. For the comparative analysis, the procedure carried out in Example 1 was chosen using oligonucleotides 1 and 2 for the polymerase chain reaction. A band having a length of about 300 bp was excised from the agarose gel for the sequence analysis. This band corresponds in length to the PTEN/MMAC1 pseudogene. After purifying the excised amplificate, Cy-5-labeled oligonucleotides with the sequence 5'- Cy-5- cag gaa aca gct atg ac - 3' were sequenced. The comparative analysis of pig liver from a butcher's with pig DNA already sequenced in this area resulted in a full correspondence. The control DNA (beef salami) has a sequence markedly differing from the pig DNA (see figure 3).

Example 3

The PCR and sequencing of exons 1 to 9 is carried out with the primers described in claims 19 to 26 with otherwise analogous reaction control. The corresponding sequences of the studied species are listed in the annex under "List of species sequences".

Example 4

This experiment is to show the species differentiation between pig DNA and human DNA using various combinations of hybridization probe pairs. The comparative analysis was made by determining the melting points of the hybridization probe pairs C1+B2, A1+B2, A1+A2, and C1+A2 in pig DNA and human DNA. The probes were chosen correspondingly and synthesized

(Tib Mol Biol, Berlin.
Probes:

A1: 5'- tgc ata ttt gtt tca tcc ggg caa att - fluorescein -3'
A2: 5'- LC Red 705 - tta aag gca caa gat ttc tat ggg ga - ph - 3'

B1: 5'- tgc ata ttt att aca tcg ggg caa att - fluorescein - 3'
B2: 5'- LC Red 640 - aag gca caa gag gcc cta gat ttc ta - ph -3'

C1: 5'- tgc ata ttt gtt aca tcg ggg taa att - fluorescein - -3'
C2: 5'- LC Red 640 - aag gca caa gag gcc cta gat ttc ta - ph -3'

A precondition for a probe hybridization is the amplification of the target area with a suitable primer pair. For this purpose, the following primer pair was chosen and synthesized:

Sense primer: PTEN se 5'- atc ttg acc aat ggc taa gtg -3' (Tib Mol Biol)
Antisense primer: Zoo44aRV 5'- ttgt ctc tgg tcc tta ctt c-3' (Amersham Pharmacia)

Both primers bind to areas of the exon 5 of the PTEN/MMAC1 gene, pseudogene and their homologues, which comprises the target regions of about 172 base pairs.
The DNAs were isolated by means of the Quiagen kit (QIAGen) from the blood of pigs and humans in accordance with the protocols of the manufacturer and stored at -20°C until they were used.
The real time PCR with hybridization probes and subsequent melting point analysis was carried out as follows: 3 rows with 4 sample batches each (glass capillary cuvette (Roche)) were provided in a cooled pipetting block (Roche). In row 1,

2 µl (50 ng) pig DNA are supplied in each batch, 2 µl (50 ng) human DNA are supplied to each batch of the 2nd row, and 2 µl distilled water are supplied in row 3 for the negative controls. The hybridization probes were used for the corresponding batches in the following combinations: 1st batch of each row: A1, 1 µl (0.1 µM) and A2, 2 µl (0.2 µM); 2nd batch of each row: C1, 1 µl (0.1 µM) and B2, 2 µl (0.2 µM); 3rd batch of each row: A1, 1 µl (0.1 µM) and B2, 2 µl (0.2 µM); 4th batch of each row: C1, 1 µl (0.1 µM) and A2, 2 µl (0.2 µM).

The following volumes and final concentrations or amounts of substrates and units of polymerase were used in all sample batches: oligonucleotides PTEN se and Zoo44aRV 2 µl (10 µM) each; $MgCl_2$ (Roche Molecular Diagnostics): 2.4 µl (4 mM); LightCycler DNA Master Hybridization Probes (Roche Molecular Diagnostics): 2 µl of a stock solution concentrated by 10 times from Taq DNA polymerase, reaction buffer, dNTP mixture and 10 mM $MgCl_2$. The final $MgCl_2$ concentration in the total reaction batch of 20 µl is 5 mM. After fully loading all reagents, the glass capillary cuvettes are closed, centrifuged at 2000 g for 1 minute and inserted in the reaction carrousel of the LightCycler provided for the capillaries.

All in all, 45 cycles are carried out in the LightCycler analysis system as follows: denaturation at 95°C for 1 second, binding of the oligonucleotides and probes at 54°C for 10 seconds and extension of the DNA strands at 72°C for 5 seconds. Denaturation was carried out at 95°C for 30 seconds before the first cycle started. The preliminary heating rate is programmed from denaturation to binding to 20°C/second, from binding to the extension of the DNA strands to 20°C/second and for the step of extension up to the denaturation to 20°C/second. The fluorescence resulting from the FRET of the probes binding complementarily side by side, is measured for observing the PCR at the end of the binding phase in each cycle. Probes which cannot bind fully to the target DNA due to sequence differences, since their binding temperature is below that of the oligonucleotides, cannot yet

be detected at that stage of fluorescence measurement.
After the complete amplification, a melting curve was finally recorded as follows: denaturation of the amplification products at 95°C for 5 seconds, cooling to 30°C with a preliminary heating rate of 20°C/second, holding of this temperature for 15 seconds and subsequent slow heating of 0.2°C/second up to 95°C. The fluorescence of the bound hybridization probes is recorded continuously up to the respective melting/dissociation during the slow temperature increase. A melting curve is recorded for each sample by recording the fluorescence signal as a function of the temperature. By forming the first negative derivative of the fluorescence against the temperature, the melting curves are converted into melting peaks.
After balancing the fluorescence signals at 640 nm and 705 nm against the standard curves prepared with a colorcompensation kit (Roche Molecular Diagnostics) for the respective wavelength, the melting points of pig DNA could be compared with those of the human DNA for the various probe combinations. In this connection, 2 melting points each could be recorded for the pig DNA for each probe pair, and one could be recorded each for the pseudogene and one for the gene. All melting points of the pig DNA differed from those of the human DNA (figure 5).

Example 5

In this experiment, the species differentiation of pig DNA from various other species of animal is shown by means of a single hybridization probe pair. For this purpose, DNAs from cattle, sheep, dwarf goat, chicken and turkey were used for the analysis (gene sequences in the annex "list of species sequences" under "sequences intron 4 exon 5").
The comparative analysis was carried out by determining the melting points of a hybridization probe pair whose sequences are specific to the pseudogene of pigs in the area of the 9 base pair deletion (A1/A2) and which was already used in Example 4. The primers used in Example 4 were employed for

the amplification. The DNA was collected from the bloods of said species by means of the QIAGen kit (QIAGen) and purified. For the real-time PCR a sample batch having 2 µl (50 ng) of the corresponding DNA each is supplied for each species and a sample batch with 2 µl distilled water is supplied for a negative control. Furthermore, 1 µl (1 µM) of the probe A1 and 2 µl (0.2 µM) of probe A2 were added by pipetting to each sample batch. To pipette further reagents the procedure of Example 1 was chosen as regards their volumes and concentrations and for the further reaction steps at the LightCycler.
A comparison of the melting points of pig DNA with those of the other animal species shows a clear distinction between pig DNA and those of other animal species (figure 6).

Example 6

In this experiment, the species differentiation between various animal species is shown by way of the following species with hybridization probes of Example 4:
Pig, deer, dog, Indian elephant, trout, quail, duck, goitred gazelle, mouse, guinea pig (gene sequences in the annex "list of species sequences" under "sequences intron4 exon5")

A comparative analysis was carried out by determining the melting points from various combinations of the hybridization probes from Example 4 in the combinations C1+B2, A1+B2, A1+A2, C1+A2, B1+B2 and B1+A2 and compiling a panel for each species of animal.

The DNA was obtained from the bloods of said species using the QIAGen kit (QIAGen) and purified. The primers listed in Example 4 were used for the amplification. For the further course of the experiment and the subsequent melting point analysis, the procedure in Example 4 was chosen, complemented by the additional probe combinations and greater numbers of species.
A panel of their melting points was compiled for the

corresponding probe combinations for each species (figure 7). A comparison of the panels of the different species yields a difference of each species from the others results as regards at least one melting point.

Figure 8 shows the mean value curves which were determined for selected probe combinations (C1+C2/A1+A2/B1+A2 and A1+C2 from 15 pig DNAs each and C1+C2/A1+A2/B1+A2 and C1+A2 from 15 human DNAs each) with the corresponding standard deviations. The experiment control corresponds to Example 4.

Furthermore, the sequences of the investigated species for exons 1-9 and the 5'-untranslated region of the PTEN/MMAC1 gene, pseudogene and their homologues are listed in the annex under "list of species sequences".

These examples for various probe combination panels can be enlarged by any number of probes and primers for the described gene PTEN/MMAC1, pseudogene and their homologues, but also for all of the other genes, and applied. A linear increase of the primers and probe combinations in the entire genome yields an exponential growth of differentiation possibilities of the most varying species among one another. By using sequence-specific probes and several and/or any number of probe combinations, a species can be distinguished from all of the other species by its very specific melting point panel.

Literature

Steck, P.A., Pershouse, M.A., Jasser, S.A., Yung, W.K.A., Lin, H., Ligon, A.H., Langford, L.A., Baumgard, M.L., Hattier, Th., Davis, T., Frye, Ch., Hu, R., Swedlund, B., Teng, D.H.F., and Tavtigian, S.V. (1997), Identification of a candidate tumor suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat. Genet. 15: 356-362

Wittwer, C.T., Herrmann, M.G., Moss, A.A., and Rasmussen, R.P. (1997), Continuous fluorescence monitoring of rapid cycle DNA amplification. BioTechniques 22: 130-138

Wittwer, C.T., Ririe, K.M., Andrew, R.V., David, D.A., Gundry, R.A. and Balis, U.J. (1997), The LightCycler™: A microvolume multisample fluorimeter with rapid temperature control. Bio Techniques: 176-181

Lay, M.J. and Wittwer, C.T. (1997), Real-time fluorescence genotyping of factor V Leiden during rapid-cycle PCR. Clinical Chemistry 43: 12; 2262-2267

Bernard, P.S., Lay, M.J. and Wittwer, C.T. (1998), Integrated amplification and detection of the C677T point mutation in the methylenetetrahydrofolate reductase gene by fluorescence resonance energy transfer and probe melting curves. Analytical biochemistry 255: 101-107

Ririe, K.M., Rasmussen, R.P. and Wittwer, C.T. (1997) Product differentiation by analysis of DNA melting curves during the Polymerase Chain reaction. Analytical Biochemistry 245: 154-160

Bernard, P.S., Pritham, G.H., and Wittwer, C.T. (1999). Color multiplexing hybridization probes using the Apolipoprotein E locus as a model system for genotyping. Analytical Biochemistry 273: 221-228

Nauck, M., Wieland, H., and März, W. (1999). Rapid, homogeneous genotying of the 4G/5G polymorphism in the promoter region of the PAI I gene by fluorescence resonance energy transfer and probe melting curves. Clinical Chemistry 45: 8; 1141-1147

Nauck, M.S., Gierens, H., Nauck, M.A., März, W. and Wieland, H. (1999), Rapid genotyping of human platelet antigen 1 (HPA-1) with fluorophore-labelled hybridization probes on the LightCycler™. British journal of Haematology 105: 803-810

Kreuzer, K.-A., Lass, U., Bohn, A., Landt, O., and Schmidt. C.A. (1999), LightCycler technology for the quantitation of brc/abl fusion transcripts. Cancer Research 59: 3171-3174

Kyger, E.M., Krevolin, M.D., and Powell, M.J. (1998), Detection of the hereditary hemochromatosis gene mutation by real-time fluorescence polymerase chain reaction and peptide nucleic acid clamping. Analytical Biochemistry 260: 142-148

Mangasser-Stephan, K., Tag, C., Reiser, A., and Gressner, A.M. (1999), Rapid genotyping of hemochromatosis gene mutations on the LightCycler with fluorescent hybridisation probes. Clinical Chemistry 45; 10: 1875-1881

Aslandis, C., Nauck, M., Schmitz and G. (1999), High-speed prothrombin G to 20210 and methylenetetrahydrofolate reductase C to 677 mutation detection using real-time fluorescence PCR and melting curves. Bio Techniques 27: 234-238

Annex

11/5/02

Brief Description of the Drawings

List - figures 1 - 8:

Figure 1: diagram of the method for identifying organisms according to Example 1

Figure 2: comparison of the determined sequences from man and African elephant according to Example 1

Figure 3: DNA sequence comparison of a pig with purchased pig liver and beef salami according to Example 2. The differences between the nucleotide sequences are shown. The 9-base pair long deletion (nucleotides 216-224) is striking in pig / pig liver.

Figure 4: positions of the hybridization probes in exon 5 of the PTEN gene, pseudogene and their homologues

Figure 5: Example 4 - melting point panels of pig and man

Figure 6: Example 5 - melting points of the probe combination A1 + A2 in pig as compared to various species

Figure 7: Example 6 - melting point panel of various species

Figure 8: standard deviations of select probes for pig (HS) and man (WT)